

Experimental Evidence for Heavy Tailed Interference in the IoT

Laurent Clavier *Senior Member, IEEE*, Troels Pedersen, Ignacio Rodriguez, Mads Lauridsen and Malcolm Egan

Abstract—5G and beyond sees an ever increasing density of connected things. As not all devices are coordinated, there are limited opportunities to mitigate interference. As such, it is crucial to characterize the interference in order to understand its impact on coding, waveform and receiver design. While a number of theoretical models have been developed for the interference statistics in communications for the IoT, there is very little experimental validation. In this paper, we address this key gap in understanding by performing statistical analysis on recent measurements in the unlicensed 863 MHz to 870 MHz band in different regions of Aalborg, Denmark. In particular, we show that the measurement data suggests the distribution of the interference power is heavy tailed, confirming predictions from theoretical models.

Index Terms—Interference, IoT, statistical models, subexponential distributions, heavy tails.

I. INTRODUCTION

As the density of connected Internet of Things (IoT) increases, interference is a pressing concern for emerging wireless networks. In addition, non-orthogonal multiple access, stringent constraints on cost, energy and computational capabilities, grant free access schemes with low coordination between devices, as well as the lack of control mechanisms mean that interference management is a key challenge. However, characterizing the interference in unlicensed bands is a non-trivial issue. One reason is the high level of heterogeneity within the network, ranging from access protocols to PHY-layer design. In particular, time-on-air, symbol duration, bandwidth and waveform selection can significantly differ from one radio access technology to another. To address this issue, a number of different probability models for the interference have been proposed sharing a key observation: *additive Gaussian noise is a poor model*.

In this paper, we analyze recent measurements in Aalborg of interference in the unlicensed European ISM 863 MHz to 870 MHz band [1], which is duty cycle limited and not based on listen before talk, such as the US 915 MHz band. Our analysis validates key features of interference statistics predicted from theoretical models. The converging variance test and analysis of the empirical survival function both indicate that the interference distribution from the measurement

data is heavy tailed. This observation is confirmed through the estimation of the tail index.

Section II sets up the notation and give the necessary background on interference models. Section III describes the expected properties of impulsive interference. Section IV studies the measurement data, states the main results and opens the discussion for the appropriate model.

II. THEORETICAL INTERFERENCE MODELS

Interference results from (partially) concurrent transmissions from different devices, called interferers, on the channel observed by a given receiver. We consider a case where a receiver is surrounded by a set of interferers, whose locations form a point process denoted as Ω (see Fig. 1). Interferer $i \in \Omega$ sits at a distance d_i from the receiver. Its signal is subject to a path loss $\ell(d_i)$ given by the positive path loss function ℓ . At a particular time-frequency (t, f) bin, the interference can be expressed as a complex number

$$X_{t,f} = \sum_{i \in \Omega} \ell(d_i) Q_{i,t,f}. \quad (1)$$

The term $Q_{i,t,f}$ includes propagation effects (e.g., multipath and shadowing) as well as the baseband emission of interferer i at time t and frequency f . In existing work, the path loss function $\ell(\cdot)$ can take different forms, but is often given by $\ell_{\gamma,\epsilon}(d) = d^{-\gamma/2} \mathbf{1}_{\{d \geq \epsilon\}}$, where γ is the path loss exponent and ϵ is the guard zone radius; that is, no interferer can be closer than ϵ from the given receiver.

A key challenge is to characterize the probability distribution of the interference, $X_{t,f}$ or the corresponding interference power $\mathcal{I}_{t,f} = |X_{t,f}|^2$. Here, we characterize $\mathcal{I}_{t,f}$ by its probability distribution function $F(x) = \mathbb{P}(X < x)$ or equivalently its survival function $S(x) = 1 - F(x)$. Analytical results have been proposed in very particular settings by using the tools of stochastic geometry. The early work by Middleton [2] showed that, if interferers were located according to a homogeneous

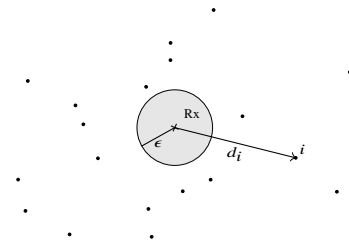


Fig. 1. Receiver (Rx) surrounded by interferers (dots). Interferer i is at distance d_i from the receiver and sits outside the guard zone of radius ϵ .

Laurent Clavier (laurent.clavier@imt-lille-douai.fr) is with IMT Lille Douai, Univ. Lille, CNRS, UMR 8520 - IEMN, France; Troels Pedersen (troels@es.aau.dk) and Ignacio Rodriguez (irl@es.aau.dk) with Dept. Electronic Systems, Aalborg University, Denmark; Mads Lauridsen (mads.lauridsen@nokia-bell-labs.com) with Nokia Bell Labs Aalborg; Malcolm Egan (malcom.egan@inria.fr) with INRIA, CITI, France. We thank COST Action CA15104 IRACON and IRCICA, USR 3380, for creating the necessary conditions for this work.

Poisson point process, non-Gaussian interference distributions naturally arise. For example, when the network radius is finite and the guard zone radius, ϵ , is non-zero, the interference distribution is known to be Middleton distributed [2], [3]. While exact, the Middleton model is difficult to work with analytically (e.g., for the purpose of receiver design), and a number of approximations have been proposed; e.g., Gaussian mixtures [4], like the ϵ -contaminated model [5].

When instead interferers are located according to a binomial point process, Weber and Andrews have shown that the resulting interference amplitude is subexponential [6]¹. In this case, the tail behaviour is dominated by the strongest interferer: F is subexponential if for independent and identically distributed (i.i.d.) X_i with common survival function S , with $x = \sum_{i=1}^n X_i$, $S(x) \rightarrow nS(X_1)$ as $x \rightarrow +\infty$.

A detailed study of the interference power has been carried out by Haenggi and Ganti [8]. In particular for interferers located according to a Poisson point process, they showed that the distribution depends heavily on the path loss attenuation coefficient γ . For example, also shown by Win and Pinto [9], in a network with infinite radius and no guard zone ($\epsilon = 0$), the interference power has the totally skewed α -stable distribution, where α depends on γ .

While these theoretical results suggest that interference is better modeled by heavy tailed distribution, there has been little experimental validation; particularly in the context of the IoT. In the following sections, we address this issue by identifying features of the Aalborg measurement data that is consistent with a family of heavy tailed distributions.

III. CLASSES OF HEAVY TAILED DISTRIBUTIONS

Several important performance metrics, such as outage probability, strongly depend on the probability that the interference is large. These rare events of high interference values play a key role, but are under-represented by the Gaussian model distribution in which the tail decay too rapidly. Alternative models such as the Middleton or α -stable with fatter tails exist.

Heavy tailed distributions are defined in general as distributions with tails that decay slower than the exponential distribution [10]. Formally, a distribution is heavy tailed if for any $M > 0$ and $t > 0$, the survival function does not satisfy

$$S(x) \leq M \exp(-tx), \quad \forall x > 0. \quad (2)$$

Failing to satisfy (2) means that the moment generating function does not exist, and hence higher order moments, the variance or even the mean are not finite.

It is natural to ask whether measurements are consistent with heavy tailed models. To clarify if such models, predicted theoretically, are indeed good approximations of the real interference behaviour, we consider two subclasses of models and their corresponding corresponding statistical tests.

A. Fat Tailed Distributions

A distribution with infinite variance is said to be fat tailed. Given a set of observations, the hypothesis that the variance

of the underlying distribution is finite can be tested using the *converging variance test* [11, Section 5.5]. Assuming that the process is a first order stationary and ergodic time-frequency process, we estimate the variance by computing the sample variance using all available time and frequency measurements (indexed by a single integer) as

$$\sigma_n^2 = \frac{1}{n} \sum_{k=1}^n \left(I_k - \bar{I}_n \right)^2, \quad (3)$$

where $\bar{I}_n = \frac{1}{n} \sum_{k=1}^n I_k$. If I has finite variance, then σ_n should rapidly converge to a finite value as n increases. If, on the other hand, the variance is infinite or very large, the convergence should not be obvious and features such as large jumps can be present as n increases.

B. Subexponential Tail Decay

A distribution F is said to have *subexponential tail decay* if there exists some $\kappa > 0$ such that its survival function satisfies

$$S(x) = x^{-\frac{1}{\kappa}} L(x), \quad \text{for } x > 0, \quad (4)$$

where κ is called the tail index and L is a slowly varying function satisfying $\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1$. Plotting $S(x)$ against $\log(x)$ yields, for subexponentially decaying F , a straight line with slope $-1/\kappa$ for x large. An exponentially decaying distribution has $\kappa = 0$ which leads to an abrupt decrease in the curve as $\log x$ increases.

Again, assuming that the interference process is first order stationary and ergodic, we estimate the marginal distribution by computing the empirical distribution function as $\hat{F}(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{I_k \leq x\}}$ and the empirical survival function as $\hat{S}(x) = 1 - \hat{F}(x)$.

C. Tail index estimation.

To quantify the results, we estimate the tail index κ . To do so, we assume the interference process is stationary and ergodic and order the observations to obtain the sequence $\{I_{(i)}\}_{i=1, \dots, n}$ such that $I_{(1)} \geq I_{(2)} \geq \dots \geq I_{(n)}$. With the k largest samples, the tail index is obtained via the Hill estimator [12]

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k \log \frac{I_{(i)}}{I_{(k)}} \quad (5)$$

This estimator is consistent, meaning that if $(k_n)_{n \in \mathbb{N}}$ is such that, as n goes to infinity, k_n goes also to infinity but k_n/n goes to zero, then $H_{k_n,n}$ tends in probability to $1/\kappa$, provided that $\{I_n\}$, besides stationary, satisfies one among a broad set of assumptions, for instance being i.i.d [12]. However, the performance of $H_{k,n}$ strongly depends on the choice of k . It is then useful to construct the Hill plot, by plotting k against $H_{k,n}$ for $k = 1, \dots, n$. The value of κ can be obtained from a flat region in the graph. To make such a plot more useful, [13] proposes to use a log scale for the x-axis, by choosing $k = \lceil n^\theta \rceil$, $0 < \theta < 1$ and plotting $H_{k,n}$ against θ . Here, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x .

¹We also note that the subexponential family has been used to model the transmit power distribution in [7].

IV. MEASUREMENT DATA ANALYSIS

We now analyze the measured interference data first reported in [1], where received power measurements were performed at five distinct locations in Aalborg (Denmark): 1) downtown shopping area; 2) a business park with office buildings; 3) hospital complex; 4) industrial area consisting of industrial production facilities and office buildings; and 5) residential area with single-family houses. At each location, measurements were performed at street level by using a radio network scanner equipped with an omni-directional antenna for a period of 2 hours. While the measurements do not reflect the interference at an elevated position, which is relevant for a base stations, they cover many IoT use cases like private homes, smart things around roads and parks, smart meters, etc... The entire on-air RF activity in the 868 MHz ISM band (863 MHz to 870 MHz) was recorded with a 7 kHz bin resolution in frequency and 200 ms sampling time yielding a sensitivity level of approximately -115 dBm. The setup and measurements are further detailed in [1].

We reduce the data by aggregating data in time-frequency windows of 200 ms and 126 kHz to fit a LoRa scheme [14]. This yields a sequence of interference samples I_1, \dots, I_n , $n = N_t \cdot N_f$ with N_t and N_f the number of time and frequency samples, respectively, as shown in Fig. 2 for a given frequency band. The presence of rare but large values can be observed.

A. Fat Tails: Converging Variance Test

We first test whether the data set is consistent with fat tails; i.e., infinite variance, or at least some infinite moments. To illustrate the variance convergence test, we plot in Fig. 3 the results on one finite variance distributions (exponential with parameter $\mu = 1$) and two subexponential distributions from the α -stable family (with $\alpha = 1.9$, slightly impulsive, and $\alpha = 1.2$, more impulsive). For the exponential model, as expected, the estimated variance converges rapidly as the sample size increases. For the two non-Gaussian α -stable distributions, there is no clear convergence even for very large sample sizes up to $n = 200\,000$. For the measured data, no clear convergence can be seen which is consistent with fat tailed models.

B. Subexponential Tail Decay

In Fig. 4, we plot the log empirical survival function versus $\log(x)$ for the different areas. To highlight the slow decay of the tails, we also plot curves corresponding to the exponential

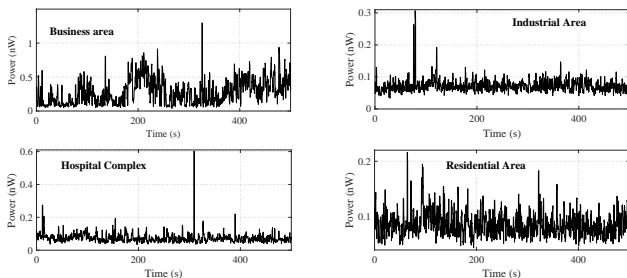


Fig. 2. Examples of recorded interference samples for different areas.

TABLE I
HILL ESTIMATE OF THE TAIL INDEX ($\theta = 0.4$).

Case	$\hat{\kappa}$
Shopping Area	1.41
Residential Area	2.34
Industrial Area	2.10
Hospital complex	0.41
Business park	1.05
Exponential ($\mu = 1$)	8.15
α -stable ($\alpha = 1.7, \beta = 1, \gamma = 0.1, \delta = 0$)	1.61

distribution ($\mu = 1$) and an α -stable distribution fitted to the data (subexponential decay). The estimated α is also given in the different plots. These visual tests clearly confirm that, on the measurement sets, interference is consistent with subexponentially decaying tails.

C. Tail Index Estimation

Fig. 5 shows the modified Hill plots. The plot for exponential distribution is decreasing while the α -stable model yields a plateau. In the measurements, the plateau is clearly present and not the monotonically decreasing line which is consistent with sub-exponential distributions. The Hill estimate of the tail index is the plateau value (see [12] for more details). Hence we read of the values at $\theta = 0.4$, see Table I. It appears that the tail indices of the data agree far better with heavy tail models (like the α -stable) than the exponential model.

V. DISCUSSION

Measurement data obtained at five different locations confirm the heavy tailed nature of the interference power distribution. The survival function and tail index are clearly those of sub-exponential random variables. Even distributions with infinite moments of order two or larger could be attractive: they can represent sudden changes in the interference that is harder to capture by distributions with all moments finite. While there is an abundance of theoretical studies of interference statistics, the measurements in Aalborg are—to the best of our knowledge—the first to clearly validate the heavy tailed nature of the interference in the context of IoT communications.

However, further measurements are required to properly explain the deviations in the tails and identify the best adapted models. Indeed, the data set is limited (one city, five locations and a receiver at the ground level). In particular, this work focused on a frequency band where users exploit random access with a duty cycle constraint (e.g., SigFox and LoRa). Other frequency bands may exploit listen-before-talk access strategies. However, at present, there are no measurement data to study the interference statistics in such cases. Also, in an IoT network, access points will more likely be at higher altitudes which could impact the interference statistics. Finally we did not analyze time or frequency dependence which is another important issue for future study.

Heavy tailed interference is known to affect the performance of signal processing and network design in wireless communication systems, and the signal processing at the receiver side

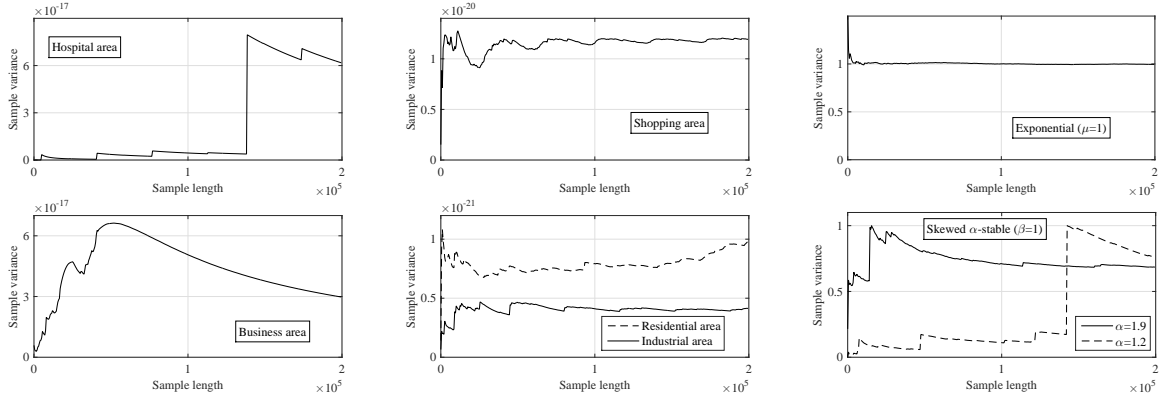


Fig. 3. Variance convergence test for four measured data sets, along with similar curves for the exponential and skewed α -stable models.

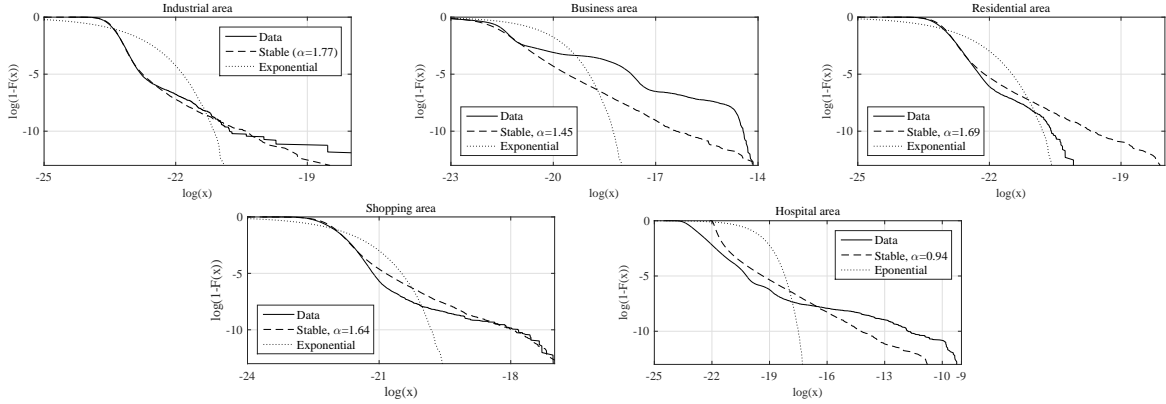


Fig. 4. Log empirical survival function $\log \hat{S}(x)$ as a function of $\log(x)$ for measured data and fitted α -stable and exponential models.

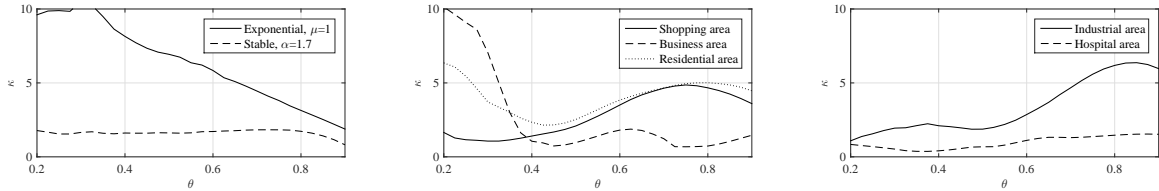


Fig. 5. Modified Hill plots for the interference data, the α -stable and exponential models. Numerical values for κ at $\theta = 0.4$ are reported in Table I.

should take it into account. Also, this work suggests several important avenues of research. One such issue is the design of channel access strategies depending on user locations to limit the impact of interference.

REFERENCES

- [1] M. Lauridsen, B. Vejlgaard, I. Z. Kovacs, H. Nguyen, and P. Mogensen, "Interference measurements in the European 868 MHz ISM Band with focus on LoRa and SigFox," in *2017 IEEE Wireless Commun. and Netw. Conf. (WCNC)*, March 2017, pp. 1–6.
- [2] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagnetic Compatibility*, vol. EMC-19, no. 3, pp. 106–127, Aug. 1977.
- [3] —, "Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1129–1149, May 1999.
- [4] N. Guney, H. Deliç, and M. Koca, "Robust detection of ultra-wideband signals in non-gaussian noise," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 4, pp. 1724–1730, Jun. 2006.
- [5] O. Alhussein, I. Ahmed, J. Liang, and S. Muhaidat, "Unified analysis of diversity reception in the presence of impulsive noise," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1408–1417, Feb. 2017.
- [6] S. Weber and J. Andrews, "Transmission capacity of wireless networks," in *Foundations and Trends in Networking*, ser. 5. NOW Publishers, 2012, no. 2-3.
- [7] A. J. Ganesh and G. L. Torrisi, "Large deviations of the interference in a wireless communication model," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3505–3517, Aug. 2008.
- [8] M. Haenggi and R. Ganti, "Interference in large wireless networks," *Foundations and Trends in Networking*, vol. 3, no. 2, pp. 127–248, 2009.
- [9] M. Win, P. Pinto, and L. Shepp, "A mathematical theory of network interference and its applications," *Proc. IEEE*, vol. 97, no. 2, pp. 205–230, Feb. 2009.
- [10] G. Peters and P. Shevchenko, *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. Hoboken, Wiley, 2015.
- [11] C. L. Nikias and M. Shao, *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, 1995.
- [12] H. Drees, L. de Haan, and S. Resnick, "How to make a hill plot," *Ann. Statist.*, vol. 28, no. 1, pp. 254–274, 02 2000. [Online]. Available: <https://doi.org/10.1214/aos/1016120372>
- [13] S. Resnick and C. Stărică, "Tail index estimation for dependent data," *The Annals of Applied Probability*, vol. 8, no. 4, pp. 1156–1183, 1998. [Online]. Available: <http://www.jstor.org/stable/2667176>
- [14] C. Pham, A. Bounceur, L. Clavier, U. Noreen, and M. Ehsan, "4 - radio channel access challenges in lora low-power wide-area networks," in *LPWAN Technologies for IoT and M2M Applications*, B. S. Chaudhari and M. Zennaro, Eds. Academic Press, 2020, pp. 65 – 102.